

Evaluating Public Sentiments of Covid-19 Vaccine Tweets Using Machine Learning Techniques

Samuel Kofi Akpatsa, Hang Lei, Xiaoyu Li and Victor-Hillary Kofi Setornyo Obeng

E-mail: samoah15@yahoo.com, hlei@uestc.edu.cn, xiaoyu33521@163.com, hillary2gh@gmail.com

School of Information and Software Engineering, University of Electronic Science and Technology of China Chengdu, Sichuan Province, 610054, P.R. China

Keywords: Natural Language Processing (NLP), Covid-19, twitter sentiment analysis, machine learning

Received: March 30, 2021

The quest to create a vaccine for covid-19 has rekindled hope for most people worldwide, with the anticipation that a vaccine breakthrough would be one step closer to the end of the deadly Covid-19. The pandemic has had a bearing on the use of Twitter as a communication medium to reach a wider audience. This study examines Covid-19 vaccine-related discussions, concerns, and Twitter-emerged sentiments about the Covid-19 vaccine rollout program. Natural Language Processing (NLP) techniques were applied to analyze Covid-19 vaccine-related tweets. Our analysis identified popular n-grams and salient themes such as "vaccine health information," "vaccine distribution and administration," "vaccine doses required for immunity," and "vaccine availability." We apply machine learning algorithms and evaluate their performance using the standard metrics, namely accuracy, precision, recall, and f1-score. Support Vector Machine (SVM) classifier proves to be the best fit on the dataset with 84.32% accuracy. The research demonstrates how Twitter data and machine learning methods can study the evolving public discussions and sentiments concerning the Covid-19 vaccine rollout program.

Povzetek: Narejena je analiza čivkov na Twitterju glede cepiv za COVID-19.

1 Introduction

The Covid-19 outbreak in late 2019 has led to tens of millions of confirmed cases and millions of deaths worldwide. The economic and social disruption of the pandemic has altered the way of life for many around the globe as public health protocols such as social distancing, wearing of masks, and travel restrictions were introduced. Although adherence to these protocols has effectively controlled the spread of the virus, there has been a global effort to develop a Covid vaccine to fight the virus head-on and help get immunity against it. Studies have indicated that at least 70% of the world's population is expected to be vaccinated to achieve herd immunity [1]. Consequently, some major pharmaceutical companies and research institutions across the globe have announced the development of Covid vaccines that promise to help ease restrictions and return the world to pre-pandemic routines.

While these developments inspire hope and optimism, other obstacles threaten the fight to eradicate the deadly virus. A significant proportion of people are unsure of the safety of the Covid vaccine, as skepticism on social media has led to the vaccine rollout exercise faced with fears, hesitancy, and opposition [2]. Meanwhile, public opinions and support for the Covid-19 vaccine are essential as this may affect whether vaccinations can be administered to large populations to achieve herd immunity.

As the Covid-19 pandemic continues to spread globally, Covid-related issues received increasing attention from the research community. Although some

studies have highlighted the socio-economic impact of the pandemic [3]–[9], analysis of Covid-19 vaccine-related issues is rare. As a social media choice for many, Twitter plays a vital role in disseminating health information in the fight against Covid-19. There is an urgent need to analyze how issues related to Covid-19 vaccine have been discussed on Twitter to understand better public perceptions, concerns, and issues that may affect their willingness to get vaccinated. Besides, identifying popular themes in tweets related to Covid-19 vaccines can play a vital role in guiding vaccine education and communication. This study examines general sentiments and opinions related to the Covid-19 vaccine rollout program by analyzing English tweets collected between January 21, 2021, and January 31, 2021. The study also identifies algorithms with suitable metrics to evaluate the performance of supervised Machine Learning classifiers on the Covid-19 vaccine tweets. The findings will be handy in assisting governments and other public health policymakers to understand trends in social media data related to the Covid-19 vaccine and make timely adjustments to vaccine education to boost public confidence in the vaccination exercise.

2 Related works

Studies on Twitter sentiment analysis provide valuable insights into real-world events and people's perceptions of these events. A review of existing literature indicates that various studies related to Twitter sentiment analysis use

popular machine learning algorithms such as Logistic Regression, Support Vector Machines, Naive Bayes to predict sentiment from tweets [10]–[14]. These algorithms essentially give better accuracy with less computational resources and are regarded as the baseline learning methods in sentiment analysis of Twitter data [15].

Recent studies demonstrate that deep learning models allow sentiment analysis systems to capture complex linguistic features and read context within a text, achieving better accuracy and performance [16]. Researchers have used these techniques to analyze unstructured data from social media posts such as Twitter [17], [18]. A related study implemented a quantum-inspired sentiment representative framework that can model semantic and the sentiment information of subjective natural language text [19]. Experimental results demonstrate the effectiveness of the framework as it significantly outperforms most state-of-the-art baselines.

Following the successful applications of these algorithms on Twitter data, an increasing number of studies have used similar approaches to analyze and understand the public response and discussions on Twitter concerning Covid-19. For example, a study analyzed the global sentiments of tweets related to Covid-19 to understand how people's sentiment in different countries has changed over time [20]. Two types of analysis were performed concerning the positive and negative sentiments, fear, and trust emotions exhibited in tweets related to *Work from Home* and *Online Learning*. The first was exploratory data analysis to provide insight into the number of daily confirmed cases. The second aspect evaluated different deep learning methods for sentiment classification on the dataset. The results showed that the general positive sentiments towards *Work from Home* and *Online Learning* have been consistently higher than negative sentiments.

A similar analysis was performed on social media posts to increase understanding of public awareness of COVID-19 pandemic trends. The research uncovers meaningful themes of concern posted by Twitter users in English during the pandemic [21]. The analyses included frequency of keywords, sentiment analysis, and topic modeling to identify and explore discussion topics over time. The results indicate that people have a negative outlook toward COVID-19.

A related study applied machine learning techniques to investigate the psychological reactions of Twitter users to Covid-19 [22]. Several salient topics were identified and categorized into themes, including "confirmed cases," "Covid-19 related death," "early signs of the outbreak," "economic impact of the pandemic," and "Preventive measures." The analysis shows that fear for the unknown nature of the coronavirus is dominant in all topics. Other successful examples of studies analyzing Twitter sentiments to determine the impact of Covid-19 on the daily aspect of life are well documented [6], [7], [9], [12], [23], [24]. These studies proved essential in assisting governments in making informed choices on managing the Covid-19 pandemic situation.

However, research on Twitter emerged sentiments on Covid-19 vaccine rollout program remain less explored in

literature. In relation to the related works, this study explores public reactions and discussions on Twitter concerning the Covid-19 vaccine rollout program. The performance comparison of different machine learning classifiers on Covid-19 vaccine Twitter dataset is also be evaluated.

3 Research methods

3.1 Research design

A purposive sampling technique was adopted in gathering Covid-19 Twitter data, published between 21st January to 31st January 2021. Our data analysis is divided into two broad parts. The first part dives into an exploratory analysis of tweets, data visualizations, and a description of the key characteristics of Covid-19 vaccine Twitter data. This approach aims to present insight and help understand public reactions and discussions on Twitter concerning Covid-19 vaccine rollout programs worldwide. The second part deals with the sentiment classification of tweets using supervised machine learning algorithms. We chose a supervised machine learning approach because our data is well labeled. The supervised learning technique allows us to measure the chosen classifiers' accuracy scores while performing sentiment classification.

3.2 Data collection

Twitter offers a variety of APIs to provide Twitter data access, including reading tweets and accessing user profiles. This study uses Twitter API and a Python script to access Covid-19 vaccine Twitter comments. A query for a hashtag (#Covid19vaccine) was run daily to collect a large number of tweet samples from around the globe. The study excluded tweets written in languages other than English. Our approach of getting the dataset is based on its availability and accessibility and how well the research community accepts the approach.

3.3 Data labeling

The labeling process aims to assign positive or negative labels to tweets. This study used human annotation to assign the value 1 (positive class) to text with positive sentiment and the value 0 (negative class) to text with negative sentiment. Some of the extracted tweets were duplicated, while others gave contradictory interpretations and proved difficult to label. As a result, some data points were removed from the dataset. The final dataset contains 15239 unique tweets, with 10519 labeled as positive and 4720 labeled as negative. The dataset contains two

	text	label
0	We have to change this The only chance we have to get through the pandemic is the COVID19Vaccine ThisIsOurShot	1
1	In my personal opinion its a civic responsibility to get the COVID19vaccine This is about not only protectin	1
2	Lack of access to patient portals could affect COVID19 vaccine for adults ages 50 and older covid19 coronavirus	0
3	They may have retired but they returned We are incredibly grateful for the many nurses that came out of retirement to help with vaccination	1
4	World Health Organization recommends using the COVID19vaccine developed by the AstraZenecaOxford even in countri	1
5	World Scientists still dont know exactly how long this immunity lasts or how strong it is though recent researc	0
6	Inaccurate information surrounding the safety of the CovidVaccines have left many women fearing the vaccinations	0
7	Essential to COVID19vaccine We need to encourage vaccination ethically so people feel comfortable and not coerced	1
8	Science really is like magic 2nd dose of my PfizerVaccine is in COVID19vaccine UofSCPharm forever grateful!	1
9	Whoop Whoop I am now 96 immune to SARS-CoV-2 the virus that causes COVID-19 thanks to the vaccine What are you waiting for	1

Figure 1: Sample tweets from the Twitter dataset.

columns (text and labels). The text column contains the text to which a label applies. These texts are transformed into features used by the model during training and prediction. The label column contains either 1 or 0, representing the sentiments of the tweet being classified. An example of the tweet dataset obtained can be seen in Figure 1.

3.4 Data preprocessing

Analyzing sentiment in tweets generally requires some fundamental cleaning and preprocessing steps to improve the quality of the dataset [25]. It includes cleaning and formatting the data before feeding it into a machine learning algorithm. Twitter datasets are often noisy with many irregularities such as punctuation marks, symbols, @links, stop words, and other special characters irrelevant for the sentiment analysis. The collected tweets are filtered using a python script to preprocess and clean the dataset to increase precision. Background noises such as white space, punctuation, hashtags, urls, special characters, hyperlinks, and stop words were removed. Also, tokenization using n-grams was applied to segment the text data and create a new document with the set n-grams, while lemmatization was applied to determine the base forms of words.

3.5 Feature extraction

This represents the extraction of lexical features such as n-grams and transforming them into a feature set that is usable by a machine learning classifier. It plays a crucial role in text classification and directly influences the text classification model [26]. Term Frequency-Inverse Document Frequency (TF-IDF) is a popular feature extraction method commonly used in text classification and sentiment analysis. TF-IDF evaluates how important a word is to a document in a dataset by converting textual representation of information into a vector space. This study applied a TF-IDF Vectorizer Python module of Scikit-learn to extract TF-IDF. First, we trained our classifier using unigram and bigram as the feature set to represent context in the Twitter data. The features are tested with the TF-IDF and trained on the classifier. After this, the trained classifier is used in predicting the test data.

3.6 Sentiment classification using machine learning

The next step after the feature extraction is to feed the feature vectors into the machine learning classifiers to perform sentiment classification. We classified the vaccine tweets using Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), and Random Forest (RF) machine learning models, and their performances were compared. Scikit-learn python library, an open-source machine learning package that provides access to machine learning classification algorithms, was used. In each experiment, the training set is used to optimize and train the machine learning algorithms, while the test set is used to evaluate the performance of the models.

3.7 Performance evaluation of machine learning classifiers

The test data was evaluated to understand better how well the classifiers performed after training. The standard metrics used to evaluate the models include accuracy, precision, recall, and f1-score. The metrics were calculated in terms of positives and negatives. The classification accuracy presents the sum of true positives and true negatives divided by the sum of all data points in the test set. Precision is the number of correctly classified positive examples divided by the total number of examples that are classified as positive. The recall measures the number of correctly classified positive examples divided by the total number of actual positive examples in the test set. The f1-score finds a balance between Precision and Recall and tell how precise and robust the classifiers were. The mathematical representation of the metrics is presented below.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives

4 Results and discussions

4.1 Principal result

We performed data analysis from the collected tweets to identify public sentiments, keyword associations, and social media trends related to the Covid-19 vaccine rollout program. We search for insights using descriptive text analysis and data visualization such as word clouds and n-gram representations. Below are brief descriptions of the data analyses on the processed Twitter dataset.

4.1.1 Word cloud representation of tweets

Word cloud was used to visualize how words are distributed across the dataset. The most recurring words provide us insight into how user sentiments about the vaccine rollout program evolved on Twitter over the study period (Figure 2).

The main goal is to examine what trend can be inferred from the word frequency in our Twitter data. The illustration from the word cloud shows that along with the search word 'covid19vaccine', words such as 'vaccine', 'dose', 'first', 'second' had many mentions. These words emphasize the awareness of the number of vaccine doses required to be fully inoculated. Names of the initially approved Covid-19 vaccines (Pfizer, Moderna, AstraZeneca) also dominated Twitter during the study period. Again, some Twitter users highlighted the crucial roles governments, healthcare professionals, and other relevant state institutions played in the vaccination

exercise as words like 'government', 'state', 'doctor', and 'hospital' were among the most frequently used words.

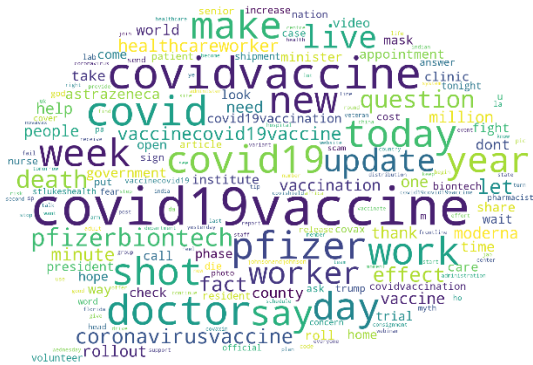


Figure 2: Word cloud representation of Covid-19 vaccine tweets.

unigram	count	bigram	count	trigram	count
0 vaccine	4048	(first, dose)	243	(received, first, dose)	59
1 httpst	1877	(side, effects)	209	(first, dose, vaccine)	46
2 get	1205	(get, vaccine)	176	(vaccine, immunity, duration)	27
3 first	1124	(vaccination, drive)	138	(vaccine, side, effects)	26
4 vaccination	992	(second, dose)	135	(worlds, largest, vaccination)	22
5 vaccines	860	(moderna, vaccine)	133	(moderna, pfizer, vaccines)	22
6 people	774	(received, first)	117	(largest, vaccination, drive)	22
7 dose	734	(get, vaccinated)	110	(dose, moderna, vaccine)	21
8 health	691	(million, doses)	102	(first, consignment, covishield)	21
9 doses	620	(pfizer, vaccine)	98	(one, step, closer)	19
10 moderna	555	(2nd, dose)	97	(received, second, dose)	18
11 pfizer	528	(johnson, johnson)	97	(first, batch, vaccine)	18
12 vaccinated	523	(healthcare, workers)	93	(received, 2nd, dose)	17
13 getting	499	(dose, vaccine)	85	(health, care, workers)	17
14 one	478	(covishield, vaccine)	81	(virtual, town, hall)	17
15 received	470	(health, workers)	79	(johnson, johnson, pauses)	16
16 new	440	(getting, vaccine)	76	(first, dose, moderna)	14
17 dr	426	(vaccine, rollout)	66	(first, dose, pfizer)	14
18 astrazeneca	351	(astrazeneca, vaccine)	66	(dose, pfizer, vaccine)	14
19 news	350	(1st, dose)	64	(second, dose, vaccine)	14

Figure 3: Top20 n-grams of Covid-19 vaccine tweets.

4.1.2 N-gram representation of tweets

N-grams are a set of consecutive words or a sequence of words in a textual document. To identify the most popular n-grams, we built a list of unique words in our Twitter data and counted each word's occurrences in a corpus. Since the Twitter dataset for this study is about Covid-19 vaccine, Covid-19 related keywords such as 'covid19vaccine', 'covidvaccine', 'covid19', 'covid', and 'coronavirus' were excluded so they do not skew our word frequency analysis. We chose uni-grams ($n=1$), bi-grams ($n=2$), and tri-grams ($n=3$) for further analysis to understand which words were used the most separately and in combination regardless of the grammar structure and semantic meaning. Figure 3 shows the most popular n-grams related to the Covid-19 vaccine tweets. These n-grams highlight how vaccine-related themes such as 'vaccine distribution,' 'vaccine administration,' and 'health engagements' dominate Twitter discussions during the study period.

From the bi-gram, phrases such as the 'first dose', 'second dose', 'receive first' imply some people have

already received their first or second dose of the vaccine. The phrase 'side effects' also gained significant recognition among Twitter users over the period. This discussion signifies the fear of potential vaccine side effects that could put the vaccination program at risk. From the tri-gram, phrases such as the 'get first dose', 'received second dose', and 'one step closer', indicate how well people have embraced the vaccination process. Also, phrases such as 'first consignment covishield', 'largest vaccination drive', and 'world largest vaccination', emphasize the global perspective of the fight against the Covid-19 pandemic. The phrase 'vaccine immunity duration' raises concern about how long Covid-19 vaccine-induced immunity will last.

4.1.3 Sentiment classification

We performed sentiment classification with four different machine learning classifiers: LR, RF, SVM, and NB, on the dataset and their performances were evaluated. To ensure that the models were learning the patterns in the data and not fitting to the noise, we implemented k-fold cross-validation technique to determine the efficiency of the classifiers. In k-fold cross-validation, the dataset is divided into k subsets which are repeated k times. For each iteration, the model is trained using k subset as the training sample and the resulting model validated on the remaining part of the data. All four different classifiers were cross-validated five (5) times and the experimental result is illustrated in Table 1.

Table 1: Cross-validation.

Runs	LR	RF	SVM	NB
1	80.75	81.38	82.79	77.01
2	81.54	82.49	81.24	78.46
3	81.25	83.26	80.23	78.90
4	81.86	81.17	82.81	77.65
5	82.33	82.41	83.74	78.14

Table 2: Evaluation of the ML Models.

Model	Accuracy	Precision	Recall	F1-score
LR	82.74	85.04	82.74	80.68
RF	83.05	84.56	83.05	81.31
SVM	84.32	84.32	84.32	83.50
NB	77.41	82.83	77.41	72.54

The result shows that the SVM classifier reaches the highest accuracy mark of 83.74 while the Naïve Bayes classifier has the lowest accuracy of 78.90 among all the classifiers (Table 1). Similarly, the predictive accuracy of the classifiers is determined to find out which model perform best in classifying the Covid-19 vaccine tweets (Table 2). The illustration in Figure 4 clearly identified SVM as the best-fit machine learning classifier on the Covid-19 vaccine Twitter dataset.

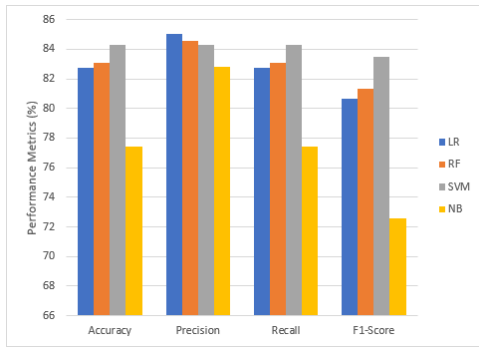


Figure 4: Performance of different ML classifiers on Covid-19 vaccine Twitter data.

4.2 Comparison to prior works

Our findings are consistent with studies using social media data to assess public responses to Covid-19. Compared with a study that implemented a Naïve Bayes model to analyze Twitter sentiments concerning Covid-19 [12], our study demonstrates that machine learning algorithms could be leveraged to study the evolving public discourse and sentiments during the Covid-19 vaccine rollout program.

Some prior Covid-related studies have described themes such as mask-wearing, social distancing, regular washing of hands, and the need for Covid vaccine as the most effective measures to stop further spread of the Covid-19 virus [4], [6]. Our study identifies a new trend of Covid-19-related discussions on Twitter during the study period. These discussions mainly focused on: (1) health information about the vaccine, (2) vaccine distribution and administration, (3) the number of vaccine doses required for immunity, and (4) questions about vaccine availability. Together with other vaccine education efforts, these themes are essential for the overall success of the vaccination program. Our n-grams were consistent with Covid-19-related studies [6], [24] that examine discussions and sentiments that emerged on Twitter and concerns about the safety measures to adopt when reopening from lockdown.

4.3 Practical implications

This study set out to examine trends that can be inferred from the targeted Twitter dataset. The study demonstrates that most of the collected tweets represent positive sentiments, which indicates the public's overall confidence in reaction to the Covid-19 vaccine rollout program. However, the n-gram representation results also suggest that a significant proportion of the public expresses negative sentiments about the Covid-19 vaccine on Twitter. There were questions about the vaccine's safety and efficacy, as the potential side effects dominate discussions over the period. Additionally, there were concerns about the vaccines' long-term protection against Covid-19. As the Covid-19 vaccine rollout program continues, more efforts from governments and relevant authorities are required to answer ongoing questions regarding vaccine choices, vaccine hesitancy, vaccine

side-effects, and the durability of the immunity response to Covid-19 vaccines.

Twitter remains a great source of information for many and can be used to explore the levels of public awareness and sentiments about the Covid-19 and its related themes. During a pandemic where people may be confined to their homes, perceptions of people about the vaccine are more likely to be inferred from social media and information online [27]. Due to the need for a worldwide Covid-19 vaccination program, understanding the threat of vaccine misinformation on social media and its negative influence on the general public's vaccine uptake is important. To encourage a positive vaccine attitude, it is suggested that social media firms devise schemes on how to promote accurate vaccine information while removing vaccine misinformation from their platforms. Besides, governments worldwide should engage their citizens on key, accurate, and timely health information regarding the vaccination program.

As evidence is still evolving, understanding the worldwide Covid-19 vaccination program's potential challenges is crucial in helping governments and relevant institutions develop schemes that will allay citizens' apprehensions about Covid-19 inoculations.

While machine learning classifiers perform relatively well on the dataset, our analysis was limited to a small collection of tweets expressed in English. Our findings may not be a true reflection of public sentiments on the vaccination program due to the risk of missing out on vital information available from tweets generated in other languages. Future work might consider the evaluation of large-scale Covid-19 vaccine Twitter datasets using deep learning models.

5 Conclusion

Our work focused on examining public discourse and reactions on Twitter concerning the Covid-19 vaccine rollout program. Popular machine learning algorithms were applied to predict sentiments from the collected tweets. Most Twitter users were optimistic during the study period, although some negative sentiments threatened the overall success of the vaccine implementation program. Also, we identified a new Covid-related discussion trend that focuses on vaccine distribution and administration, the number of vaccine doses required for immunity, and other health information about the vaccine. These findings can be a handy tool to help policymakers and relevant authorities anticipate the appropriate measures that can be taken to mitigate any potential challenges to the vaccine rollout program. The pressing need to achieve herd immunity against Covid-19 requires timely reactions to address the concerns of the general public to boost trust and confidence in the vaccination program.

Data availability

The model, code, and the dataset are available in the GitHub repository: <https://github.com/askasnr/Covid-19-vaccine-tweets-Dataset>

Acknowledgements

This study was supported by the National Key R&D Program of China, Grant No. 2018YFA0306703.

Conflict of interest

The authors declare that the research was conducted without any commercial or financial relationships that could be interpreted as a potential conflict of interest.

References

- [1] R. Aguas, R. M. Corder, J. G. King, G. Goncalves, M. U. Ferreira, and M. G. M. Gomes, “Herd immunity thresholds for SARS-CoV-2 estimated from unfolding epidemics,” medRxiv, 2020. <https://doi.org/10.1101/2020.07.23.20160762>
- [2] S. Chawla and M. Mehrotra, “Impact of emotions in social media content diffusion,” Informatica, vol. 45, no. 6, 2021. <https://doi.org/10.31449/inf.v45i6.3575>
- [3] M. Cinelli et al., “The covid-19 social media infodemic,” Sci. Rep., vol. 10, no. 1, pp. 1–10, 2020. Cross-reference
- [4] W.-Y. S. Chou and A. Budenz, “Considering Emotion in COVID-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence,” Health Commun., vol. 35, no. 14, pp. 1718–1722, 2020. <https://doi.org/10.1080/10410236.2020.1838096>
- [5] E. M. Martey, H. Lei, X. Li, and O. Appiah, “Effective Image Representation using Double Colour Histograms for Content-Based Image Retrieval,” Informatica, vol. 45, no. 7, 2021. <https://doi.org/10.31449/inf.v45i7.3715>
- [6] J. Xue et al., “Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach,” J. Med. Internet Res., vol. 22, no. 11, p. e20550, 2020. <https://doi.org/10.2196/20550>
- [7] J. Samuel, G. G. Ali, M. Rahman, E. Esawi, Y. Samuel, and others, “Covid-19 public sentiment insights and machine learning for tweets classification,” Information, vol. 11, no. 6, p. 314, 2020. <https://doi.org/10.3390/info11060314>
- [8] Addo Prince Clement et al., “COVID-19 and Actor Well-being: A Serial Mediated Moderation of Mask Usage and Personal Health Engagement,” Asian J. Immunol., vol. 5, pp. 44–53, 2021. <https://doi.org/10.1080/21645515.2021.2008729>
- [9] A. D. Dubey, “Twitter Sentiment Analysis during COVID19 Outbreak,” Available SSRN 3572023, 2020. <http://dx.doi.org/10.2139/ssrn.3572023>
- [10] F. H. Rachman and others, “Twitter Sentiment Analysis of Covid-19 Using Term Weighting TF-IDF And Logistic Regression,” in 2020 6th Information Technology International Seminar (ITIS), 2020, pp. 238–242. <https://doi.org/10.1109/itis50118.2020.9320958>
- [11] S. Naz, A. Sharan, and N. Malik, “Sentiment classification on twitter data using support vector machine,” in 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), 2018, pp. 676–679. <https://doi.org/10.1109/wi.2018.00-13>
- [12] K. H. Manguri, R. N. Ramadhan, and P. R. M. Amin, “Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks,” Kurdistan J. Appl. Res., pp. 54–65, 2020. Cross-reference
- [13] D. Suleiman, W. Etaawi, and A. Awajan, “Recurrent Neural Network Techniques: Emphasis on Use in Neural Machine Translation,” Informatica, vol. 45, no. 7, 2021. <https://doi.org/10.31449/inf.v45i7.3743>
- [14] W. Etaawi, D. Suleiman, and A. Awajan, “Deep Learning Based Techniques for Sentiment Analysis: A Survey,” Informatica, vol. 45, no. 7, 2021. <https://doi.org/10.31449/inf.v45i7.3674>
- [15] S. Elbagir and J. Yang, “Sentiment Analysis of Twitter Data Using Machine Learning Techniques and Scikit-learn,” in Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence, 2018, pp. 1–5. <https://doi.org/10.1145/3302425.3302492>
- [16] S. K. Akpatsa, X. Li, and H. Lei, “A Survey and Future Perspectives of Hybrid Deep Learning Models for Text Classification,” in International Conference on Artificial Intelligence and Security, 2021, pp. 358–369. https://doi.org/10.1007/978-3-030-78609-0_31
- [17] H. Jelodar, Y. Wang, R. Otji, and H. Huang, “Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach,” arXiv Prepr. arXiv2004.11695, 2020. <https://doi.org/10.1109/JBHL.2020.3001216>
- [18] A. Roy and M. Ojha, “Twitter sentiment analysis using deep learning models,” in 2020 IEEE 17th India Council International Conference (INDICON), 2020, pp. 1–6. <https://doi.org/10.1109/indicon49873.2020.9342279>
- [19] Y. Zhang, D. Song, P. Zhang, X. Li, and P. Wang, “A quantum-inspired sentiment representation model for twitter sentiment analysis,” Appl. Intell., vol. 49, no. 8, pp. 3093–3108, 2019. Cross-reference
- [20] M. Mansoor, K. Gurumurthy, V. R. Prasad, and others, “Global Sentiment Analysis Of COVID-19 Tweets Over Time,” arXiv Prepr. arXiv2010.14234, 2020. <https://doi.org/10.48550/arXiv.2010.14234>
- [21] S. Boon-Itt and Y. Skunkan, “Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study,” JMIR Public Heal. Surveill., vol. 6, no. 4, p. e21978, 2020. <https://doi.org/10.2196/21978>

- [22] J. Xue, J. Chen, C. Chen, C. Zheng, S. Li, and T. Zhu, “Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter,” *PLoS One*, vol. 15, no. 9, p. e0239441, 2020.
<https://doi.org/10.1371/journal.pone.0239441>
- [23] P. C. Addo, F. Jiaming, N. B. Kulbo, and L. Liangqiang, “COVID-19: fear appeal favoring purchase behavior towards personal protective equipment,” *Serv. Ind. J.*, vol. 40, no. 7–8, pp. 471–490, Jun. 2020.
<https://doi.org/10.1080/02642069.2020.1751823>
- [24] M. E. Ahmed, M. R. I. Rabin, and F. N. Chowdhury, “COVID-19: Social media sentiment analysis on reopening,” *arXiv Prepr. arXiv2006.00804*, 2020.
<https://doi.org/10.48550/arXiv.2006.00804>
- [25] Y. HaCohen-Kerner, D. Miller, and Y. Yigal, “The influence of preprocessing on text classification using a bag-of-words representation,” *PLoS One*, vol. 15, no. 5, p. e0232525, 2020.
<https://doi.org/10.1371/journal.pone.0232525>
- [26] V. Singh, B. Kumar, and T. Patnaik, “Feature extraction techniques for handwritten text in various scripts: a survey,” *Int. J. Soft Comput. Eng.*, vol. 3, no. 1, pp. 238–241, 2013. Cross-reference
- [27] W. H. Organization and others, “Behavioural considerations for acceptance and uptake of COVID-19 vaccines: WHO technical advisory group on behavioural insights and sciences for health, meeting report, 15 October 2020,” 2020.
<https://iris.who.int/handle/10665/337335>

